DOCUMENT RESUME

ED 388 681 TM 023 714

AUTHOR Kramer, Gene A.
TITLE A Cross-validation of the Design Components Influencing the Difficulty of Orthographic-Projection Spatial-Ability Items.
PUB DATE Apr 95
NOTE 17p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Difficulty Level; *Item Response Theory; *Spatial Ability; Test Construction; *Test Items
IDENTIFIERS Cross Validation; Dental Admission Testing Program; Orthographic Structure; Projection (Psychology); Rasch Model; Variance (Statistical)

ABSTRACT

The present study is designed to cross-validate the findings of an earlier component analysis of orthographic-projection, spatial-ability items. The earlier research identified four design components that contribute to the difficulty of orthographic-projection items. The research found that increasing Rasch item difficulties on component frequencies results in the components accounting for 66 percent (N=10) of the variance in item difficulty. Because of the small number of items in the analysis, however, the explained percentage of variance may have been a function of the limited degrees of freedom. Using an enlarged item pool of 19 items from the Perceptual Ability Test of the Dental Admission Test battery, the present research findings show that a similar set of design components accounts for 25 percent (N=17) of the variance in item difficulty. (Contains 5 tables and 20 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A Cross-validation of the Design Components influencing the Difficulty of Orthographic-projection Spatial-ability Items

ED 388 681

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GENE A. KRAMER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Gene A. Kramer
American Dental Association

Paper presented at the 1995 Meeting of the
National Council on Measurement in Education
San Francisco

BEST COPY AVAILABLE

Abstract

The present study is designed to cross-validate the findings of an earlier component analysis of orthographic-projection, spatial-ability items. The earlier research identified four design components that contribute to the difficulty of orthographic-projection items. The research found that regressing Rasch item difficulties on component frequencies resulted in the components accounting for 66 percent (N=10) of the variance in item difficulty. Because of the small number of items in the analysis, however, the explained percentage of variance may have been a function of the limited degrees of freedom. Using an enlarged item pool, the present research findings show that a similar set of design components account for 25 percent (N=17) of the variance in item difficulty.

A Cross-validation of the Design Components influencing the Difficulty of Orthographic-projection Spatial-ability Items

Component analysis is a method for specifying the contribution of factors to the difficulty of test items. There are two distinct purposes for conducting this type of analysis. First, component analysis can be used to validate constructs (Stenner, Smith, & Burdick, 1983) and to confirm theories of cognitive processing. For example, Embretson and Wetzel (1987) examined the relative impact of cognitive variables such as encoding, coherence processing, and text mapping on paragraph comprehension items. Second, component analysis can be used to identify design components that influence item difficulty (Smith, Kramer, & Kubiak, 1992). Information regarding such components allows those individuals involved in test construction to develop items with associated difficulties that are either targeted at the ability level of a group of examinees or are targeted across a full range of examinee abilities.

Spatial-ability items are particularly amenable to component analysis. It is a straight-forward task to identify and quantify components that tend to contribute to item difficulty. Examples of specific components are 'number of surfaces in a figure', 'visualization of hidden parts', and 'mental reorientation of figures in space'.

The objective of the present study is to confirm those design components that contribute to the difficulty of orthographic-projection, spatial-ability items. Previous research findings (Kramer, 1992) showed that four components accounted for 66 percent

(multiple R=0.81, n=10) of the variability in the difficulties of the items. However, because the number of items was small relative to the number of components in the regression analysis, it was hypothesized that the percentage of explained variance found may have been artificially high.

Research conducted on other item types has found a range of results. For paper-folding items, the components accounted for 86 percent of the variance (multiple R=0.93, n=45)(Smith & Green, 1985; Smith, Kramer, & Kubiak, 1992). For aperture-type items, the components accounted for 57 percent (multiple R=0.76, n=26)(Kramer, 1993). Further, components accounted for 45 percent of the variance (multiple R=0.67, n=79) in the difficulty of cube-type items (Kramer, Smith, & Kubiak, 1989), and components accounted for 39 percent of the variance (multiple R=0.62, n=30) in the difficulty of form-development items (Kramer & Smith, 1990).

The orthographic-projection item is more complex than the paper-folding, aperture, and cube-type items in that there are a variety of figure shapes and orthographic projections. In this regard, the orthographic projection is similar to the form-development item. The identification of the components of the items that contribute to difficulty is less trivial in the case of orthographic-projection and form-development items because of this complexity.

In the earlier analysis of the orthographic-projection items, identification of the components of difficulty focused on components similar to those identified in studies involving other types of items. These components are subsumed by three global

factors. For the global factor of complexity, these included the 'number of surfaces in the figure' and the 'number of surfaces in the correct projection'. For orientation, the component concerned whether the top or end view is involved in correctly responding to the item. For visualization, the component is the 'number of hidden surfaces in the projection'.

Instrument

The items used in this study were taken from the Perceptual Ability Test (PAT), which is part of the Dental Admission Test battery (DAT). This test battery is used for graduate admission to all dental schools in the United States and is taken by approximately 8,500 persons each year. The test is administered twice each year in Fall and Spring. The PAT consists of 75 scored and 15 experimental multiple-choice items that are divided into six item types, the orthographic-projection items being one of these item types. There are 15 orthographic-projection items on current editions of the test. The remaining item types are angle discriminations, paper-foldings, apertures, form developments, and cubes (Kramer, Kubiak, & Smith, 1989).

An orthographic-projection item is shown in Figure 1. The examinee is required to visualize one of three views (top, front, end) of a three-dimensional object when given the other two views. One way of solving this type of spatial-ability problem is for the examinees to mentally construct the three-dimensional figure based on the two views provided in the item stem. Then the examinees are to reorient the figure and visualize the view that they are to select from among the alternatives. While solving some spatial-

ability problems might be facilitated by logical analysis, orthographic projections would seem to depend primarily on the spatial ability of the examinees.

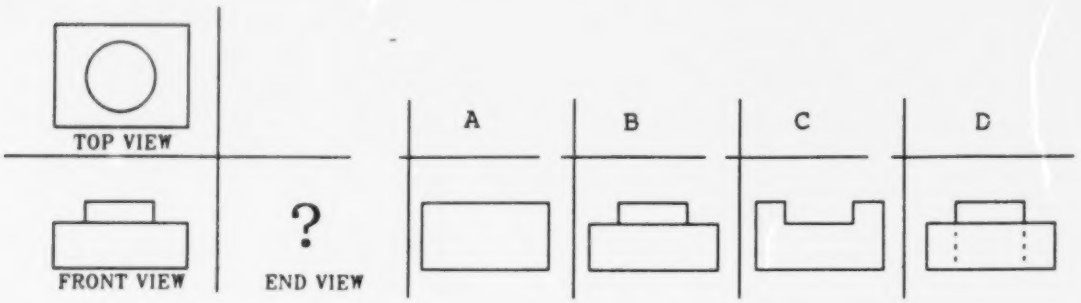

Figure 1. A sample orthographic projection-type, spatial-ability item.

Method

While a number of different approaches can be employed in component analysis, linear regression has been successfully used in many studies (Green & Smith, 1987; Kramer, Smith, & Kubiak, 1989; Smith & Kramer, 1988). With this approach, it is necessary to identify those components of the item that appear to contribute to the item difficulty. The components are specified by their absence or presence in the item or by the frequency with which they occur in the item. Component analysis then involves estimating logit item difficulties, investigating the fit of the items to the measurement model, and regressing the logit item difficulties on the previously specified component frequencies. For the purposes of this research, the Rasch measurement model was used.

The Rasch ability scale is based on the log odds of responding to an item correctly. The metric is defined in terms of these log-odds or logits. The Rasch model is defined by the expression:

$$P(x_{vi}=1 \mid \beta_v, \delta_i) = \frac{\exp(\beta_v - \delta_i)}{[1+\exp(\beta_v - \delta_i)]}$$

where P_{vi} is the probability of person v correctly responding to item i, β_v is the ability of the person v, and δ_i is the difficulty of the item i (Wright & Stone, 1979).

The underlying log-ability scale is centered on zero and typically ranges from a -5.00 to a +5.00 logits, with more negative values indicating relatively easier items and lower-scoring examinees. In like manner, more positive values indicate relatively more difficult items and higher-scoring examinees.

The data used in this study were taken from the Spring 1991 and Spring 1994 administrations of the Dental Admission Test. The Spring 1991 edition includes 12 items, and the Spring 1994 edition includes 15 items. Eight items appear on both forms of the test. Random samples of 2,475 and 3,564 examinees were selected from the 1991 and 1992 administrations, respectively. The two test forms were calibrated using BIGSTEPS (Wright & Linacre, 1992). Missing values were considered incorrect. The Spring 1991 item-by-person response matrix was submitted to BIGSTEPS. The resulting difficulties for the eight common items were used as anchor values in calibrating the Spring 1994 responses. In this way, all 19 items were placed on a common logit metric.

The fit of the responses to the Rasch measurement model was assessed through use of the total fit statistics (Wright & Stone, 1979). The unweighted total fit statistic, the outfit, is used to assess the fit of the item responses to the Rasch measurement

model; it has an expected value of 0.0 with a standard deviation of 1.0 (Smith, 1989). The difficulty estimates were then regressed on the frequency of the components using a stepwise regression to provide estimates of the amount of variance in the item difficulties that is explained by variance in the components (Fischer, 1973; Mislevy, 1981; Embretson & Wetzel, 1987; Green & Smith, 1987; SAS, 1993, 1985).

To ensure that the items were measuring the same underlying trait, the 1991 and 1994 item-by-person response matrices were factor analyzed separately (SAS, 1993, 1985). Principal component analyses were used. Items not loading on the primary factors were removed, and the data were reanalyzed.

Results and Conclusions

The design components incorporated in this study were based on the earlier research (Kramer, 1992). One additional component, i.e., 'hidden surfaces in figure', was included in the present study in an attempt to capture a greater proportion of the variance in the item difficulties. Table 1 shows the five components and their frequency of occurrence in the 19 items.

The results of the calibration indicate that the data fit the Rasch model fairly well, considering the fact that the test was not designed to fit the Rasch model. Of the 19 items in the calibrations, all showed fit statistics comparable to the other item types on the test. Therefore, given the sample size, there appears to be no reason to believe that these data were inappropriate for analysis with the Rasch model. Unusually high fit statistics were associated with two of the items, however.

Table 1

Frequency of Difficulty Components

Component	Range
Surfaces in figure	7 - 27
Surfaces in orthographic projection	1 - 5
Hidden Surfaces in Figure*	1 - 21
Hidden Surfaces in orthographic projection	5 - 22
Orientation**	0 - 1

* Not included in (Kramer, 1992).
** End view is coded 0; top view is coded 1.

The stepwise regression analysis of the 19 items resulted in a multiple R of 0.39 for four components. This multiple R is lower than the 0.81 found in the earlier research. This would tend to confirm that the earlier result was artificially high as a function of the degrees of freedom. The summary of the results are shown in Table 2. Note that one component did not enter the equation; the F-statistic reflecting the contribution of the component 'hidden surfaces in orthographic projection' to the regression model was not significant at the 0.99 level.

Tables 3 and 4 show the results of the principal component analyses of individual items included in the Spring 1991 and Spring 1994 data sets. Table 3 provides the eigenvalues and variances explained by the first five factors that emerged from the analyses. As shown, the factor eigenvalues decline sharply between the first and second factors and then are relatively stable across the remaining factors. As discussed by Tabachnick and Fidell (1983), this shows that, overall, the item sets are evaluating one underlying trait. Table 4 provides the factor patterns for the first two factors extracted by the analyses. As shown, only one

Table 2

Regression Summary

Component	R^2	F	Prob > F
Surfaces in Projection	0.09	1.58	0.226
Hidden Surfaces in Figure	0.10	0.33	0.577
Orientation	0.12	0.23	0.639
Surfaces in Figure	0.15	0.62	0.446
Multiple R = 0.39			

Number of items = 19

item in each data set loads on the second factor rather than on the first as do the other items in the data sets. Moreover, the outfit statistics associated with these items were high relative to the statistics shown for the other items in the data set. For both items, the outfit was +9.9. The outfit statistics for the remaining items did not exceed $\pm$7.0.

The results of the separate principal component analyses of the Spring 1991 and Spring 1994 item-by-person response matrices showed that all but two items loaded on the same factor. This indicated that these two items were evaluating a trait different than the other items. Because of this, these two items were removed from the data, and the reduced data set was analyzed using stepwise regression. The results are shown in Table 5.

The multiple R of 0.50 found for orthographic projections is not surprising. This value is lower than those found for other types of spatial-ability items because the other types are less complex at least in part because the shapes of the figures are more regular. Because of this greater complexity, it is more difficult to identify design components that capture significant amounts of

Table 3

Eigenvalues and Explained Variances for Factors 1 through 5

	Factor				
	1	2	3	4	5
Spring 1991					
Eigenvalue	3.27	1.10	1.02	0.85	0.83
Explained Variance	0.27	0.09	0.09	0.07	0.07
Spring 1994					
Eigenvalue	4.81	1.09	0.89	0.83	0.79
Explained Variance	0.32	0.07	0.06	0.06	0.05

variance in the difficulty of items. Spatial-ability items with more regular shapes are not necessarily less complex, however.

The order of entry of the design components in the stepwise regression equation and the standardized regression coefficients provide useful insights into the structural characteristics of this item type. Of the four components entering the equation, two are related to problem complexity -- 'surfaces in projection' and 'surfaces in figure'. The regression coefficient for the component 'surfaces in figure' is positive. This suggests that as the spatial problem becomes more complex, the item difficulty increases. The regression coefficient for the 'orientation' is positive as well. This suggests that certain views may increase the difficulty. The regression coefficients for the remaining two components are negative. The coefficients, however, are close to 0.00, and, as such, the sign is of little practical significance. The component 'hidden surfaces in projection' did not enter the equation.

Table 4

Principal Component Analyses of Orthographic-projection Items: Factor Patterns

Spring 1991		
Factor	1	2
ITEM.14	0.54863	0.00402
ITEM.15	0.60997	-0.17588
ITEM.16	0.54086	0.01231
ITEM.17	0.51621	0.26246
ITEM.18	0.48261	-0.30347
ITEM.19	0.53830	-0.21964
ITEM.20	0.64552	-0.21435
ITEM.21	0.57756	0.11127
ITEM.22	0.49395	-0.23880
ITEM.23	0.42562	0.45883
ITEM.24	0.24009	0.72115
ITEM.25	0.53441	0.09591
Spring 1994		
Factor	1	2
ITEM.16	0.55740	0.31340
ITEM.17	0.53032	0.32286
ITEM.18	0.55457	0.01408
ITEM.19	0.54707	0.16039
ITEM.20	0.63236	0.06607
ITEM.21	0.39214	0.50071
ITEM.22	0.60130	-0.01534
ITEM.23	0.47312	-0.40207
ITEM.24	0.61100	-0.12446
ITEM.25	0.49454	-0.36256
ITEM.26	0.61779	0.21015
ITEM.27	0.54935	-0.47756
ITEM.28	0.59007	-0.10979
ITEM.29	0.62109	-0.08490
ITEM.30	0.66116	0.03817

This pattern of findings is consistent with the findings of the previous study as well as the findings of research on other item types. The components are related to three global factors, i.e., complexity, visualization, and reorientation. The components 'surfaces in figure' and 'surfaces in orthographic projection' are complexity factors, while the component 'hidden surfaces in figure' is related to the amount of information that has to be visualized.

Table 5

Regression Summary

Component	R^2	F	Prob > F
Surfaces in Projection	0.15	2.71	0.226
Orientation	0.17	0.30	0.577
Surfaces in Figure	0.18	0.20	0.639
Hidden Surfaces in Figure	0.25	1.09	0.317
Multiple R = 0.50			

Number of items = 17

The view that must be visualized is related to the global factor of orientation.

Despite the complexity of the orthographic-projection item type, the same basic ideas about the underlying components of item structure and the relationship of these characteristics to the cognitive processes involved in the solution of the items emerged from the analysis. Including additional components in the analysis would undoubtedly capture a greater proportion of the variance in the item difficulties. However, the limited degrees of freedom preclude the inclusion of such additional components even though they may be closely related to the global factors.

With regard to using this type of approach, it is interesting to note that a clear determination of the relative importance of the design components that contribute to the difficulty of items can be problematic in some cases. Removing items from the regression analysis that do not load on the same factor as the other items results in a change in the order of importance. As discussed in Green, Kluever, and Wright (1994), the item design components are multicollinear. Because of this, item dependent

variances in the component values have a strong influence on the regression results. Therefore, a slight change in the variances that might follow from removing items from the data, leads to design components entering the regression equation in a different order. It seems clear that the best estimate of the proportion of variance accounted for by the component values and the order of importance of the components results from the removal of those items that do not seem to fit the definition of the underlying variable. However, the change in the ordering introduces uncertainty to the findings.

While this type of analysis is problematic in some regards, it remains a powerful tool in understanding the structural characteristics of items of different types. As further understanding of the relationships emerge, these analyses should lead to the ability to design items that directly test theories about the underlying cognitive processes involved as well as facilitate item-construction activities. Further application of this methodology to more typical multiple-choice achievement topics should be possible and can lead to a fuller understanding of the underlying variables measured by these achievement tests.

References

Embretson, S. E. & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement, 11,* 175-193.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 359-374.

Green, K. E., Kluever, R. C., & Wright, B. (1994). Predicting item difficulties from item characteristics. *Rasch Measurement Transactions, 8,* 354-355.

Green, K.E. & Smith, R.M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics, 12,* 369-381.

Kramer, G. A. (1993, April). *A cross-validation of the design components influencing the difficulty of aperture-type, spatial-ability items.* Paper presented at the meeting of the National Council on Measurement in Education, Atlanta.

Kramer, G. A. (1992, April). *An investigation of components influencing the difficulty of orthographic-projection, spatial-ability items.* Paper presented at the meeting of the National Council on Measurement in Education, San Francisco.

Kramer, G.A., Kubiak, A.T., & Smith R.M. (1989). Construct and predictive validity of the Perceptual Ability Test. *Journal of Dental Education, 53,* 119-125.

Kramer, G.A., & Smith, R.M. (1990, April). *An investigation of components influencing the difficulty of form-development items.* Paper presented at the annual meeting of the National Council on Measurement in Education, Boston.

Kramer, G.A., Smith, R.M., & Kubiak, A.K. (1989, March). *A cross-validation of components influencing the difficulty of cube items.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Mislevy, R. (1981). A general linear model for the analysis of Rasch item threshold estimates. Unpublished dissertation. University of Chicago.

SAS Companion for the Microsoft Windows Environment (1993). Cary, North Carolina: SAS Institute Inc.

SAS User's Guide: Statistics (1985). Cary, North Carolina: SAS Institute Inc.

Smith, R.M. (1989, March). *Item and person fit in the Rasch model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Smith, R.M. & Green, K.E. (1985). *Components of difficulty in paper folding items.* A paper presented at the annual meeting of the American Educational Research Association, Chicago.

Smith, R.M. & Kramer, G.A. (1988, April). *Component analysis of factors influencing the difficulty of perceptual ability items.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Smith, R. M., Kramer, G. A., & Kubiak, A. T. (1992). Components of difficulty in spatial ability items. In Mark Wilson (Ed.), *Objective measurement: Theory into practice.* Norwood, NJ: Ablex Publishing Corporation.

Stenner, A.J., Smith, M. & Burdick, D.S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement, 20,* 305-316.

Tabachnick, B.G., & Fidell, L.S. (1983). *Using multivariate statistics.* New York: Harper & Row.

Wright, B.D., & Linacre, J.M. (1992). A user's guide to BIGSTEPS: Rasch-model computer program, Version 2.2. Chicago.

Wright, B.D. & Stone, M. (1979). *Best test design.* Chicago: MESA Press.

Correspondence about this paper and requests for copies should be sent to Gene A. Kramer, Department of Testing Services, American Dental Association, 211 East Chicago Avenue, Suite 1840, Chicago, Illinois 60611-2678.